

# ПОИСК НЕЯВНЫХ СВЯЗЕЙ КЛАССОВ СООБЩЕСТВ В СОЦИАЛЬНЫХ СЕТЯХ ПОСРЕДСТВОМ ГРАФОВОЙ ВИЗУАЛИЗАЦИИ

А.Ю. Кайда, А.О. Савельев

Томский политехнический университет

E-mail: ayk13@tpu.ru

## Введение

Большая часть данных в социальных сетях является общедоступной, поэтому работу с такими данными можно осуществлять без специального режима доступа. Несмотря на это, на основе открытых данных возможно получение неявных выводов о взаимосвязи между различными пользователями/сообществами, ценных для социологических или кроссдисциплинарных исследований.

Целью работы является построение социального графа, задачей которого является отображение неявных связей между различными классами сообществ социальной сети (на примере сети ВКонтакте).

## Описание подхода

Социальный граф – графическое представление данных о пользователях/группах пользователей, собранных из социальной сети, в виде набора вершин и связывающих их ребер. В качестве вершин могут выступать пользователи или сообщества, в то время как ребро отображает тот или иной тип связи (ее наличие или отсутствие) согласно собранным данным [1].

Гипотеза исследования заключается в том, что различные классы сообществ в социальной сети, не имея очевидных ссылок друг на друга, могут быть тесно связаны друг с другом через недостающие звенья – ссылки первого порядка («сообщество-ссылка-сообщество»). В случае, если количество таких сообществ существенно, сообщества-ссылки могут относиться к сообществам, схожим с сообществами из контрольного списка и принадлежать к одному из классов сообществ. В таком случае, использование предлагаемого подхода возможно также для пополнения контрольных списков новыми данными (сообществами) [2].

В качестве узлов графа задаются сообщества, где цветовая разметка указывает на принадлежность того или иного сообщества из контрольного списка к определенному классу из заданных. Отдельным цветом помечаются узлы на те сообщества, которые являются ссылками. Далее отрисовываются ребра между узлами («сообщество из контрольного списка» – «сообщество по ссылке»). В случае, если ссылка из одного сообщества контрольного списка указывает на аналогичное другое сообщество из контрольного списка, ребро рисуется между двумя сообществами из контрольного списка, ни один из двух узлов не помечается как сообщество-ссылка.

Ожидается, что на графе будут выявлены кластеры связанных сообществ одного класса, а также возможно возникновение связей с другими классами. Поскольку направленность связи не имеет значения, граф является неориентированным [3].

## Графовая визуализация данных

Входными данными для проведения исследования являются данные о группах, собранные в файл в формате .xlsx, т.е. в файл, данные в котором представлены в форме, доступной для чтения пользователем. Входной файл содержит наименования сообществ, ссылки на сообщества и классы, к которым отнесены данные сообщества.

Далее была осуществлена серия дополнительных запросов к социальной сети посредством программного интерфейса приложения (API) [3] для получения дополнительных данных (ссылок на другие сообщества, указанные в сообществах контрольного списка). Данный интерфейс позволяет получить только те данные, которые являются открытыми для просмотра любого пользователя социальной сети (или гостя). Выходные данные запросов проходят фильтрацию и передаются на визуализацию. Визуализация данных организуется с помощью библиотек для языка программирования Python [4,5], предназначенных для построения двумерных изображений – графиков и схем – и отрисовки элементов графа.

В исходных данных были представлены ссылки на 98 ресурсов социальной сети ВКонтакте, относящимся к 23-м различным категориям. В ходе исследований было выявлено, что данные по части сообществ невозможно получить, т.к. сообщество или перестало существовать (сменило адрес, было официально заблокировано по запросу или закрылось по иным причинам), или ссылка относилась к ресурсу иного типа, не являющегося сообществом. Итоговая выборка состояла из 87 сообществ, рассортированных по 8 классам. Результаты апробации гипотезы представлены на рисунке 1.

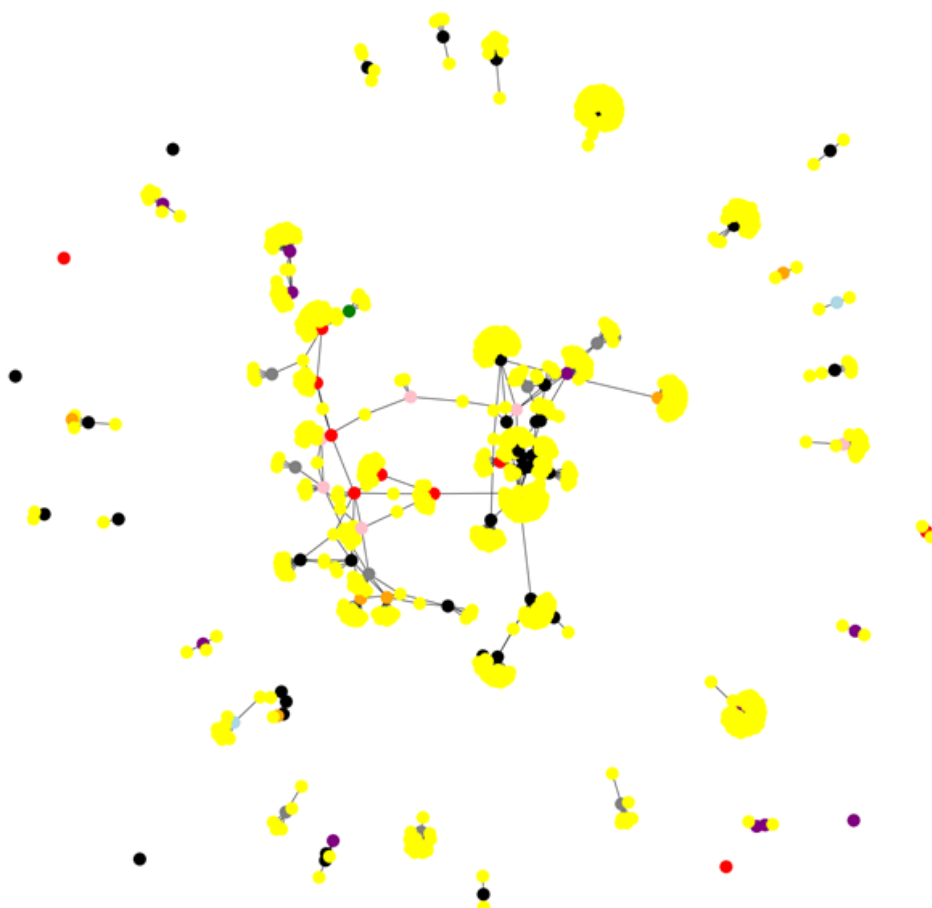


Рис. 1. Социальный граф, отображающий связи между классами сообществ

### Заключение

В результате апробации было выявлено два крупных кластера сообществ, в каждом из кластеров была выявлена тесная связь сообществ друг с другом (в том числе связи типа «сообщество-сообщество»), при этом также была выявлена связь данных кластеров друг с другом. При этом также были выявлены связи с сообществами других классов через ссылки первого порядка. Ожидалось, что так или иначе большая часть сообществ контрольного списка так или иначе связана друг с другом.

Результаты, представленные на полученном графе, подтверждают применимость данного подхода и его перспективность для будущих исследований. Дополнительным критерием может являться интенсивность (важность) связи, выражаемая графически в различной толщине ребра в зависимости от величины исследуемого параметра. Для более детализированных исследований требуется увеличение объема выборки (от 1000 сообществ).

### Список использованных источников

1. Guo, T., Luo, J., Dong, K., Yang, M. Differentially private graph-link analysis based social recommendation. *Information Sciences*, 463-464, 2018. – pp. 214–226.

2. Gaillard, P., Aupetit, M., & Govaert, G. (2008). Learning topology of a labeled data set with the supervised generative Gaussian graph. *Neurocomputing*, 71(7-9), 2008. – 1283–1299.
3. Опе О. Теория графов. — М.: Наука, 1968. — 336 с.
4. Список методов VK API | Разработчикам [Электронный ресурс]. — URL: <https://vk.com/dev/methods> (дата обращения 20.11.2019).
5. Matplotlib: Python plotting — Matplotlib 3.1.2 documentation [Электронный ресурс]. — URL: <https://matplotlib.org/> (дата обращения 20.11.2019).
6. NetworkX — NetworkX [Электронный ресурс]. — URL: <https://networkx.github.io/> (дата обращения 20.11.2019).